D.J
#6 3-29-02

862.C2175

Priority Papers

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:)
: Examiner: Not Assigned
YASUO OKUTANI, ET AL.)
: Group Art Unit: 2644
Application No.: 09/818,581)
:
Filed: March 28, 2001)
:
For: SPEECH SYNTHESIS APPARATUS AND) July 2, 2001
METHOD, AND STORAGE MEDIUM :

Commissioner for Patents
Washington, D.C. 20231

CLAIM TO PRIORITY

Sir:

Applicants hereby claim priority under the International Convention and all rights to which they are entitled under 35 U.S.C. § 119 based upon the following Japanese Priority Application:

JAPAN	2000-099420	March 31, 2000.

A certified copy, with the translation of the first page of the priority document, is enclosed.

Applicants' undersigned attorney may be reached in our Washington, D.C. office by telephone at (202) 530-1010. All correspondence should continue to be directed to our address given below.

Respectfully submitted,

Attorney for Applicants

Gary M. Jacobs
Registration No. 28,861

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200
GMJ\cmv

09/818,581
Yoshio OKUTANI, et al
3-28-07

(translation of the front page of the priority document of Japanese Patent Application No. 2000-099420)

PATENT OFFICE
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: March 31, 2000

Application Number : Patent Application 2000-099420
Applicant(s) : Canon Kabushiki Kaisha

April 20 2001

Commissioner,
Patent Office Kouzo OIKAWA

Certification Number 2001-3033156

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２０００年　３月３１日

出願番号
Application Number: 特願２０００－０９９４２０

出願人
Applicant(s): キヤノン株式会社

２００１年　４月２０日

特許庁長官
Commissioner,
Japan Patent Office 及川耕造

出証番号　出証特２００１－３０３３１５６

【書類名】　特許願
【整理番号】　4172011
【提出日】　平成12年 3月31日
【あて先】　特許庁長官殿
【国際特許分類】　G01L 5/04
【発明の名称】　音声情報処理装置とその方法と記憶媒体
【請求項の数】　21
【発明者】
【住所又は居所】　東京都大田区下丸子３丁目３０番２号　キヤノン株式会社内
【氏名】　奥谷　泰夫
【発明者】
【住所又は居所】　東京都大田区下丸子３丁目３０番２号　キヤノン株式会社内
【氏名】　小森　康弘
【特許出願人】
【識別番号】　000001007
【氏名又は名称】　キヤノン株式会社
【代理人】
【識別番号】　100076428
【弁理士】
【氏名又は名称】　大塚　康徳
【電話番号】　03-5276-3241
【選任した代理人】
【識別番号】　100101306
【弁理士】
【氏名又は名称】　丸山　幸雄
【電話番号】　03-5276-3241

【選任した代理人】

【識別番号】　100115071

【弁理士】

【氏名又は名称】　大塚　康弘

【電話番号】　03-5276-3241

【手数料の表示】

【予納台帳番号】　003458

【納付金額】　21;000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　0001010

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　音声情報処理装置とその方法と記憶媒体

【特許請求の範囲】

【請求項１】　音素素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力手段と、

音声合成に使用する素片辞書に登録する音声素片を、前記歪出力手段から出力された歪に基づいて選択する素片登録手段と、

を有することを特徴とする音声情報処理装置。

【請求項２】　前記歪出力手段は、前記音声素片を他の音声素片と接続することによって生じる接続歪と前記音声素片を変形することによって生じる変形歪とに基づいて、前記歪を求めることを特徴とする請求項１に記載の音声情報処理装置。

【請求項３】　テキストデータを入力するテキスト入力手段と、

前記入力されたテキストデータの言語解析を行なう言語解析手段と、

前記言語解析手段による解析結果に基づいて前記所定の韻律情報を生成する韻律生成手段を更に有することを特徴とする請求項１又は２に記載の音声情報処理装置。

【請求項４】　前記接続歪及び変形歪により決定される歪を基準として音声素片系列のＮbest系列を求めるＮbest決定手段を更に有し、

前記素片登録手段は、前記音声素片系列のＮbest系列を基に前記素片辞書に登録する音声素片を選択することを特徴とする請求項２又は３に記載の音声情報処理装置。

【請求項５】　前記素片登録手段は、前記接続歪と前記変形歪との重み付き加算に基づいて、前記素片辞書に登録する音声素片を選択することを特徴とする請求項２又は３に記載の音声情報処理装置。

【請求項６】　前記歪出力手段は、各音声素片のケプストラム距離を用いて前記接続歪を決定することを特徴とする請求項２乃至５のいずれか１項に記載の音声情報処理装置。

【請求項７】　前記歪出力手段は、変形前の音声素片と変形後の音声素片に

おけるケプストラム距離を用いて前記変形歪を決定することを特徴とする請求項２乃至５のいずれか１項に記載の音声情報処理装置。

【請求項８】　前記歪出力手段は、前記変形歪を記憶したテーブルを有し、当該テーブルを参照して前記変形歪を決定することを特徴とする請求項２乃至５のいずれか１項に記載の音声情報処理装置。

【請求項９】　前記歪出力手段は、前記接続歪を記憶したテーブルを有し、当該テーブルを参照して前記接続歪を決定することを特徴とする請求項２乃至５のいずれか１項に記載の音声情報処理装置。

【請求項１０】　前記素片辞書を用いてテキストデータを音声合成する音声合成手段を更に有することを特徴とする１乃至９のいずれか１項に記載の音声情報処理装置。

【請求項１１】　音素素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力工程と、

音声合成に使用する素片辞書に登録する音声素片を、前記歪出力工程で出力された歪に基づいて選択する素片登録工程と、

を有することを特徴とする音声情報処理方法。

【請求項１２】　前記歪出力工程は、前記音声素片を他の音声素片と接続することによって生じる接続歪と前記音声素片を変形することによって生じる変形歪とに基づいて、前記歪を求めることを特徴とする請求項１１に記載の音声情報処理方法。

【請求項１３】　テキストデータを入力するテキスト入力工程と、

前記入力されたテキストデータの言語解析を行なう言語解析工程と、

前記言語解析工程による解析結果に基づいて前記所定の韻律情報を生成する韻律生成工程を更に有することを特徴とする請求項１１又は１２に記載の音声情報処理方法。

【請求項１４】　前記接続歪及び変形歪により決定される歪を基準として音声素片系列のＮbest系列を求めるＮbest決定工程を更に有し、

前記素片登録工程では、前記Ｎbest系列を基に前記素片辞書に登録する音声素片を選択することを特徴とする請求項１２又は１３に記載の音声情報処理方法。

【請求項１５】　前記素片登録工程では、前記接続歪と前記変形歪との重み付き加算に基づいて、前記素片辞書に登録する音声素片を選択することを特徴とする請求項１２又は１３に記載の音声情報処理方法。

【請求項１６】　前記歪出力工程では、各音声素片のケプストラム距離を用いて前記接続歪を決定することを特徴とする請求項１２乃至１５のいずれか１項に記載の音声情報処理方法。

【請求項１７】　前記歪出力工程では、変形前の音声素片と変形後の音声素片におけるケプストラム距離として変形歪を定量化して決定することを特徴とする請求項１２乃至１５のいずれか１項に記載の音声情報処理方法。

【請求項１８】　前記歪出力工程では、前記変形歪を記憶したテーブルを有し、当該テーブルを参照して前記変形歪を決定することを特徴とする請求項１２乃至１５のいずれか１項に記載の音声情報処理方法。

【請求項１９】　前記歪出力工程では、前記接続歪を示すテーブルを有し、当該テーブルを参照して前記接続歪を決定することを特徴とする請求項１２乃至１５のいずれか１項に記載の音声情報処理方法。

【請求項２０】　前記素片辞書を用いてテキストデータを音声合成する音声合成工程を更に有することを特徴とする１１乃至１９のいずれか１項に記載の音声情報処理方法。

【請求項２１】　請求項１１乃至２０のいずれか１項に記載の方法を実行するプログラムを記憶したことを特徴とする、コンピュータにより読取り可能な記憶媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、音声合成で使用される素片辞書を作成する音声情報処理装置及びその方法と記憶媒体に関するものである。

【０００２】

【従来の技術】

近年、音声素片を１ピッチ波形単位で複製及び、或いは削除しながら所望のピ

ッチ間隔で貼り合わせて編集し（PSOLA：ピッチ同期波形重畳法）、それらの音声素片を接続して音声合成する音声合成方法が主流となっている。

【０００３】

【発明が解決しようとする課題】

このような技術を利用して音声合成された音声には、音声素片を編集することによる歪（以下、変形歪）と、音声素片を接続することによる歪（以下、接続歪）とが含まれる。これら２つの歪が、合成された音声の品質劣化を引き起こす大きな要因となる。中でも、素片辞書に登録できる音声素片の数が制限される状況下では、音声合成時に、このような歪が小さくなるように音声素片を選択する余地がほとんど残されていない場合がある。特に、一つの音韻環境について１つの音声素片しか素片辞書に登録できない場合には、歪が小さくなるように音声素片を選択する余地は全くなく、このような素片辞書を用いると、変形歪や接続歪による合成音声の品質劣化は避けられないものとなる。

【０００４】

本発明は上記従来例に鑑みてなされたもので、接続歪や変形歪に基づき歪の影響を考慮して、素片辞書に登録する音声素片を選択することによって音声合成の音質劣化を抑制する音声情報処理装置及びその方法と記憶媒体を提供することを目的とする。

【０００５】

【課題を解決するための手段】

上記目的を達成するために本発明の音声情報処理装置は以下のような構成を備える。即ち、

音素素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力手段と、

音声合成に使用する素片辞書に登録する音声素片を、前記歪出力手段から出力された歪に基づいて選択する素片登録手段と、

を有することを特徴とする。

【０００６】

また前記歪出力手段は、前記音声素片を他の音声素片と接続することによって

生じる接続歪と前記音声素片を変形することによって生じる変形歪とに基づいて、前記歪を求めることを特徴とする。

【０００７】

上記目的を達成するために本発明の音声情報処理方法は以下のような工程を備える。即ち、

音素素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力工程と、

音声合成に使用する素片辞書に登録する音声素片を、前記歪出力工程で出力された歪に基づいて選択する素片登録工程と、

を有することを特徴とする。

【０００８】

【発明の実施の形態】

以下、添付図面を参照して本発明の好適な実施の形態を詳細に説明する。

【０００９】

［実施の形態１］

図１は、本発明の実施の形態に係る音声情報処理装置のハードウェア構成を示すブロック図である。尚、本実施の形態では、一般的なパーソナルコンピュータを音声合成装置として用いる場合について説明するが、本発明は専用の音声情報処理装置であっても、また他の形態の装置であっても良い。

【００１０】

図１において、１０１は制御メモリ（ＲＯＭ）で、中央処理装置（ＣＰＵ）１０２で使用される各種制御データを記憶している。ＣＰＵ１０２は、ＲＡＭ１０３に記憶された制御プログラムを実行して、この装置全体の動作を制御している。１０３はメモリ（ＲＡＭ）で、ＣＰＵ１０２による各種制御処理の実行時、ワークエリアとして使用されて各種データを一時的に保存するとともに、ＣＰＵ１０２による各種処理の実行時、外部記憶装置１０４から制御プログラムをロードして記憶している。この外部記憶装置は、例えばハードディスク、ＣＤ－ＲＯＭ等を含んでいる。１０５はＤ／Ａ変換器で、音声信号を示すデジタルデータが入力されると、これをアナログ信号に変換してスピーカ１０９に出力して音声を再

生する。１０６は入力部で、オペレータにより操作される、例えばキーボードや、マウス等のポインティングデバイスを備えている。１０７は表示部で、例えばＣＲＴや液晶等の表示器を有している。１０８はバスで、これら各部を接続している。１１０は音声合成ユニットである。

【００１１】

以上の構成において、本実施の形態の音声合成ユニット１１０を制御するための制御プログラムは外部記憶装置１０４からロードされてＲＡＭ１０３に記憶され、その制御プログラムで用いる各種データは、制御メモリ１０１に記憶されている。これらのデータは、中央処理装置１０２の制御の下にバス１０８を通じて適宜メモリ１０３に取り込まれ、中央処理装置１０２による制御処理で使用される。Ｄ／Ａ変換器１０５は、制御プログラムを実行することによって作成される音声波形データ（ディジタル信号）をアナログ信号に変換してスピーカ１０９に出力する。

【００１２】

図２は、本実施の形態に係る音声合成ユニット１１０のモジュール構成を示すブロック図で、この音声合成ユニット１１０は、大きく分けて、素片辞書２０６に音声素片を登録するための処理を実行する素片辞書作成モジュールと、テキストデータを入力し、そのテキストデータに対応する音声を合成して出力する処理を行なう音声合成モジュールの２つのモジュールを有している。

【００１３】

図２において、２０１は、入力部１０６又は外部記憶装置１０４から任意のテキストデータを入力するテキスト入力部、２０２は解析辞書、２０３は言語解析部、２０４は韻律生成規則保持部、２０５は韻律生成部、２０６は素片辞書、２０７は音声素片選択部、２０８は音声素片編集・接続部、２０９は音声波形出力部、２１０は音声データベース、２１１は素片辞書作成部、２１２はテキストコーパスである。このテキストコーパス２１２には、入力部１０６などを介して種々の内容のテキストを入力することができる。

【００１４】

まず、音声合成モジュールについて説明する。この音声合成モジュールでは、

言語解析部２０３が、解析辞書２０２を参照して、テキスト入力部２０１から入力されるテキストの言語解析を行なう。こうして解析された結果が韻律生成部２０５に入力される。韻律生成部２０５は、言語解析部２０３における解析結果と、韻律生成規則保持部２０４に保持されている韻律生成規則に関する情報とを基に、音韻系列と韻律情報を生成して音声素片選択部２０７及び音声素片編集・接続部２０８に出力する。続いて、音声素片選択部２０７は、韻律生成部２０５から入力される韻律生成結果を用いて、素片辞書２０６に保持されている音声素片から対応する音声素片を選択する。音声素片編集・接続部２０８は、韻律生成部２０５から入力される韻律生成結果に従って、音声素片選択部２０７から出力される音声素片を編集及び接続して音声波形を生成する。こうして生成された音声波形は、音声波形出力部２０９で出力される。

【００１５】

次に、素片辞書作成モジュールについて説明する。

【００１６】

このモジュールでは、素片辞書作成部２１１が、後述する手順に基づいて、音声データベース２１０の中から音声素片を選び出して素片辞書２０６に登録する。

【００１７】

次に、上記構成を備えた本実施の形態の音声合成処理について説明する。

【００１８】

図３は、図２の音声合成モジュールにおける音声合成処理（オンライン処理）の流れを示すフローチャートである。

【００１９】

まずステップＳ３０１で、テキスト入力部２０１は、文、文節、単語等の単位毎にテキストデータを入力してステップＳ３０２に移る。ステップＳ３０２では、言語解析部２０３により当該テキストデータの言語解析を行う。次にステップＳ３０３に進み、音韻生成部２０５はステップＳ３０２で解析された結果と所定の韻律規則とに基づいて、音韻系列と韻律情報を生成する。次にステップＳ３０４に進み、各音韻毎にステップＳ３０３で得られた韻律情報と所定の音韻環境と

に基づいて、音声素片選択部２０７が素片辞書２０６に登録されている音声素片を選択する。次にステップＳ３０５に進み、その選択された音声素片及びステップＳ３０３で生成された韻律情報とに基づいて、音声素片編集・接続部２０８により音声素片の編集および接続を行なってステップＳ３０６に進む。ステップＳ３０６では、音声素片編集・接続部２０８によって生成された音声波形を、音声波形出力部２０９が音声信号として出力する。このようにして、入力されたテキストに対応する音声が出力されることになる。

【００２０】

図４は、図２で示した素片辞書作成モジュールの、より詳細な構成を示すブロック図で、前述の図２と共通する部分は同じ番号で示し、かつ本実施の形態の特徴である素片辞書作成部２１１の構成をより詳細に示している。

【００２１】

図４において、４０１はテキスト入力部、４０２は言語解析部、４０３は解析辞書、４０４は韻律生成規則保持部、４０５は韻律生成部、４０６は音声素片検索部、４０７は音声素片保持部、４０８は音声素片編集部、４０９は変形歪決定部、４１０は接続歪決定部、４１１は歪決定部、４１２は歪保持部、４１３はNbest決定部、４１４はNbest保持部、４１５は登録素片決定部、４１６は登録素片保持部である。

【００２２】

以下、詳しく説明する。

【００２３】

テキスト入力部４０１は、テキストコーパス２１２から、例えば文単位にテキストデータを取り出して言語解析部４０２に出力する。言語解析部４０２は、解析辞書４０３を参照してテキスト入力部４０１から入力されたテキストデータを解析する。韻律生成部４０５は、言語解析部４０２で解析された解析結果に基づいて音韻系列を生成し、韻律生成規則保持部４０４が保持する韻律生成規則（アクセントパターン、自然降下成分、ピッチパターン等）を参照して韻律情報を生成する。音声素片検索部４０６は、韻律生成部４０５で生成される韻律情報と音韻系列とに従って音声データベース２１０から、各音韻毎に、所定の音韻環境を

考慮した音声素片を検索する。こうして検索された音声素片は一旦、音声素片保持部４０７に保持される。音声素片編集部４０８は、韻律生成部４０５で生成された韻律情報に合わせて音声素片保持部４０７に保持されている音声素片を編集する。この編集には、韻律情報に合わせて音声素片同士を接続する処理や、またその音声素片同士の接続に際して音声素片の一部を削除する等して変形する処理などが含まれる。

【００２４】

変形歪決定部４０９は、各音声素片の変形前と変形後の音響的特徴の変化から変形歪を決定する。接続歪決定部４１０は、音韻系列において一つ前の音声素片の終端付近の音響的特徴と当該音声素片の始端付近の音響的特徴から、これら音声素片同士が接続された場合の接続歪を決定する。歪決定部４１１は、変形歪決定部４０９で決定された変形歪と、接続歪決定部４１０で決定された接続歪とを考慮し、音韻系列ごとにトータルの歪（歪値ともいう）を決定する。歪保持部４１２は、歪決定部４１１で決定された各音声素片に至る歪の値を保持する。Ｎbest決定部４１３は、Ａ*（エースター）探索アルゴリズムを用いて、音韻系列毎に歪が最小となる上位Ｎ通りの最適パスを求める。Ｎbest保持部４１４は、Ｎbest決定部４１３で求めたＮ個の最適パスを入力テキストごとに保持する。登録素片決定部４１５は、Ｎbest保持部４１４に保持されている、各音韻ごとにＮbestの結果から、その頻度順に、素片辞書２０６に登録する音声素片を選び出す。登録素片保持部４１６は、登録素片決定部４１５により選ばれた音声素片を保持する。

【００２５】

図５は、図４で示す素片辞書作成モジュールにおける処理の流れを示すフローチャートである。

【００２６】

まずステップＳ５０１で、テキスト入力部４０１がテキストコーパス２１２から一文ずつテキストデータを取り出す。取り出せるテキストデータが存在しなくなると、最終的に登録する音声素片を決定するステップＳ５１２に進む。テキストデータが存在する場合はステップＳ５０２に進み、言語解析部４０２において

、解析辞書４０３を使って、その入力されたテキストデータの言語解析を行なってステップＳ５０３に進む。ステップＳ５０３では、韻律生成部４０５により、韻律生成規則保持部４０４が保持する韻律生成規則と、ステップＳ５０２における言語解析結果とに基づいて韻律情報並びに音韻系列を生成する。次にステップＳ５０４に進み、ステップＳ５０３で生成された音韻系列内の各音韻を順次処理する。このステップＳ５０４で未処理の音韻が存在しない場合はステップＳ５１１に進むが、未処理の音韻が存在する場合はステップＳ５０５に進む。ステップＳ５０５において、音声素片検索部４０６は、各音韻毎に音韻環境及び韻律規則を満足する音声素片を音声データベース２１０から検索して音声素片保持部４０７に保存する。

【００２７】

例えば具体例で説明すると、テキストデータとして「こんにちわ」が入力されると、それが言語解析され、アクセントやイントネーション等を含む韻律情報が生成される。そして、この「こんにちわ」は、例えばｄｉｐｈｏｎｅを音韻の単位として用いた場合、以下のような音韻系列に分解される。

【００２８】

こ　　ん　　に　　ち　　わ

/k　k.o　o.X　X.n　n.i　i.t　t.i　i.w　w.a　a/

なお、ここで「X」は、音声「ん」を示し、「/」は無声音を示す。

【００２９】

次にステップＳ５０６に進み、その検索された複数の音声素片について順次処理する。未処理の音声素片が存在しない場合はステップＳ５０４に戻って次の音韻の処理に進むが、存在する場合はステップＳ５０７に進んで、現在の音韻の音声素片を処理する。ステップＳ５０７では、音声素片編集部４０８が、上述の音声合成処理時と同じ手法を用いて音声素片の編集を行なう。ここでいう音声素片の編集とは、例えばピッチ同期波形重畳法（PSOLA）などの処理である。この音声素片の編集には、その音声素片と韻律情報を用いる。音声素片の編集が終了したらステップＳ５０８に進み、変形歪決定部４０９により、現在の音声素片の変形前と変形後における音響的特徴の変化を変形歪として算出する（この詳細は後

述する）。次にステップＳ５０９に進み、接続歪決定部４１０により、現在の音声素片とその一つ前の音韻の音声素片の全てとの接続歪を算出する（この処理についても詳しく後述する）。次にステップＳ５１０に進み、歪決定部４１１は、変形歪と接続歪から現在の音声素片に至るパスの全てについて歪値を決定する（後述する）。そして現在の音声素片に至るパスの歪値の上位Ｎ個（Ｎ：求めたいＮbestの個数）と、そのパスを表わす一つ前の音韻の音声素片へのポインタを歪保持部４１２に保持してステップＳ５０６に戻り、現在の音韻において未処理の音声素片が存在するかどうかを調べる。

【００３０】

こうしてステップＳ５０６で、各音韻における全ての音声素片が処理され、更にステップＳ５０４で全ての音韻が処理されるとステップＳ５１１に進む。ステップＳ５１１において、Ｎbest決定部４１３は、Ａ*探索アルゴリズムを用いたＮbest探索を行ない、上位Ｎ位までの最適パス（音声素片系列ともいう）を求め、これをＮbest保持部４１４に保持してステップＳ５０１に戻る。

【００３１】

こうして全テキストに対する処理が終了するとステップＳ５０１からステップＳ５１２に進み、登録素片決定部４１５は、音韻ごとに全テキストのＮbest結果に基づいて所定の頻度の高い以上を選択して音声素片を素片辞書２０６に登録する。尚、このＮbestにおけるＮの値は、予備実験などから経験的に与えておく。こうして決定された音声素片は、登録素片保持部４１６を介して素片辞書２０６に登録される。

【００３２】

図６は、本実施の形態に係る図５のステップＳ５０８における変形歪の求め方を説明する図である。

【００３３】

ここでは、PSOLA法によりピッチ間隔を広げる場合について図示している。矢印はピッチマーク、点線は変形前と変形後のピッチ素片の対応関係を表わしている。本実施の形態では、各ピッチ素片（微細素片ともいう）の変形前後のケプストラム距離に基づいて変形歪を表わす。具体的には、まず変形後のあるピッチ素

片（例えば６０で示す）のピッチマーク６１を中心にハニング窓６２（窓長２５．６ミリ秒）をかけ、そのピッチ素片６０を周辺のピッチ素片を含めて切り出す。こうして切り出したピッチ素片６０をケプストラム分析する。次に、ピッチマーク６１に対応する変形前のピッチ素片６３のピッチマーク６４を中心にして同じ窓長のハニング窓６５でピッチ素片を切り出し、変形後の場合と同様にしてケプストラムを求める。このようにして求めたケプストラム同士の距離を、着目しているピッチ素片６０の変形歪として、変形後のピッチ素片とそれに対応する変形前のピッチ素片間の変形歪の総和をPSOLAで採用されるピッチ素片数Ｎｐで割った値を、その音声素片の変形歪とする。こうして求められる変形歪を式で記述すると以下のようになる。

【００３４】

$$Dt = \Sigma\Sigma \mid Corg\ i,j - Ctar\ i,j \mid / Np$$

ここで最初のΣは、ｉ＝１からＮまでの総和を示し、次のΣはｊ＝０～１６までの総和を示している。またＣtar i,jは、変形後のｉ番目のピッチ素片のケプストラムのｊ次元目の要素を表わし、同様に、Ｃorg i,jは、変形後に対応する変形前のピッチ素片のケプストラムのｊ次元目の要素を表わしている。

【００３５】

図７は、本実施の形態における接続歪の求め方を説明する図である。

【００３６】

この接続歪は、一つ前の音韻の音声素片と現在の音声素片との接続箇所において生じる歪を示し、ここではケプストラム距離を用いて表わす。具体的には、音声素片境界が存在するフレーム７０，７１（フレーム長５ミリ秒、分析窓幅２５．６ミリ秒）と、それを挟む前後それぞれ２フレームからなる計５フレームを接続歪の算出対象としている。ここでケプストラムは、０次（パワー）～１６次（パワー）までの計１７次元ベクトルとする。そして、このケプストラムベクトルの各要素の差の絶対値の和を、現在注目している音声素片における接続歪とする。即ち、図７の７００で示すように、一つ前の音韻の音声素片における終端部のケプストラムベクトルの各要素をＣpre i,j（ｉ：フレーム番号、フレーム番号の“０”が音声素片境界があるフレームを示し、ｊがベクトルの要素番号を示す

）とする。また、図７の７０１で示すように、注目音声素片における始端部のケプストラムベクトルの各要素をＣcur i,jとすると、現在注目している音声素片の接続歪Ｄcは、

$$Dc=\Sigma\Sigma\,|\,Cpre\ i,j\ -\ Ccur\ i,j\,|$$

となる。ここで最初のΣはｉ＝－２～２の総和を、次のΣはｊ＝０～１６までの総和を示す。

【００３７】

図８は、本実施の形態に係る歪決定部４１１による、音声素片における歪の決定過程を図示したものである。本実施の形態において、音韻単位はdiphone（ダイフォン）とする。

【００３８】

図中、一つの円がある音韻における１つの音声素片を示し、円内の数字は、この音声素片に至る歪値の総和の最小値を示している。また四角で囲まれた数字は、一つ前の音韻の音声素片と現在注目している音韻の音声素片との間の歪値を示している。また矢印は、現在注目している音韻の音声素片と一つ前の音韻の音声素片との関連を示している。ここでは説明のため、ｎ番目の音韻（現在注目している音韻）のｍ番目の音声素片をＰn,mとする。この音声素片Ｐn,mの最も小さい歪値から上位Ｎ個（Ｎ：求めたいＮbestの数）までに対応する音声素片を一つ前の音韻の中から取り出し、その中のｋ番目の歪値をＤn,m,kとし、その歪値に対応するの一つ前の音韻の音声素片をＰＲＥn,m,kとすると、ＰＲＥn,m,kを介して音声素片Ｐn,mに至るパスにおける歪値の総和Ｓn,m,kは、

$$Sn,m,k\ =\ Sn-1,x,0\ +\ Dn,m,k\qquad(\text{但し、}\ x=PREn,m,k)$$

となる。

【００３９】

本実施の形態における歪値について説明する。本実施の形態では歪値Ｄtotal（上記説明におけるＤn,m,kに相当する）を、上述の接続歪Ｄcと変形歪Ｄtの重み付き和として定義する。

【００４０】

$$Dtotal\ =\ w\times Dc\ +\ (1-w)\times Dt\qquad:(0\leqq w\leqq 1)$$

ここで重み係数ｗは、予備実験など経験的に求められる係数で、ｗ＝０の場合は、歪値が変形歪Ｄｔのみで説明され、ｗ＝１の場合は、歪値が接続歪Ｄｃのみに依存することになる。

【００４１】

歪保持部４１２では、各音韻の音声素片Ｐｎ,ｍ毎に、上位Ｎ個の歪値Ｄｎ,ｍ,ｋと、それらに対応する一つ前の音韻の音声素片ＰＲＥｎ,ｍ,ｋと、ＰＲＥｎ,ｍ,ｋを介してＤｎ,ｍ,ｋに至るパスの歪値の総和Ｓｎ,ｍ,ｋをそれぞれ保持する。

【００４２】

図８では、現在注目している音声素片Ｐｎ,ｍに至るパスの総和の最小値が「２２２」となる例を示す。この時の音声素片Ｐｎ,ｍの歪値は、Ｄｎ,ｍ,1(k=1)であり、この歪値Ｄｎ,ｍ,1に対応する一つ前の音韻の音声素片は、ＰＲＥｎ,ｍ,1（図８のＰｎ-1,ｍ８１に相当する）である。８０は、音声素片ＰＲＥｎ,ｍ,1と音声素片Ｐｎ,ｍとを接続するパスである。

【００４３】

図９は、Ｎbestの決定過程を図示したものである。

【００４４】

ステップＳ５１０の終了時点で、各音声素片において、上位Ｎ個の情報がそれぞれ求まっている（フォワード探索）。Ｎbest決定部４１３では、音韻系列の末尾の音声素片９０から逆順に枝を伸ばしながらＮbestパスを求める（バックワード探索）。この枝を伸ばすノードの選択は、予測値（線の横の数字）とそこに至る総歪値の和（歪値は四角の中の数字で示される）が最小となるものである。ここでいう予測値とは、音声素片Ｐｎ,ｍにおけるフォワード探索結果の最小歪Ｓｎ,ｍ,0に相当する。この場合、予測値と実際に左端までに至る最小パスの歪が等しいので、Ａ*探索アルゴリズムの性質により最適パスが求まることが保証される。

【００４５】

図９は、第１位の最適パスが決定された状態を示す図である。

【００４６】

図中、丸が音声素片を示し、その丸の中の数字が歪み予測値、太い実線が第一位のパス、四角の中の数字が歪値、線の横の数字が予測歪み値を示している。次

に第２位のパスを求めるために、二重丸のノードの中で、予測値とそこに至る総歪値の和が最小となるノードを選択し、それに繋がる一つ前の音韻の音声素片の全て（最大Ｎ個）に枝を伸ばす。この伸ばした先のノードが二重丸で表現されている。この操作を繰り返すことにより、上位Ｎ個のパスが総歪値の順に決定される。この図９は、Ｎ＝２として枝を伸ばした場合の例を示す図である。

【００４７】

このようにして本実施の形態１によれば、歪の最も小さいパスを形成する音声素片を選択して、それを素片辞書に登録することができる。

【００４８】

［実施の形態２］

前述の実施の形態１では、音韻の単位としてdiphoneを用いる場合について記述したが、本発明はこれに限定されるものではなく、音素や半diphoneなどを単位としてもよい。半diphoneとは、diphoneを音素境界で２つに分割したもののことである。この半diphoneを単位とした場合のメリットについて簡単に説明する。任意のテキストを合成する場合、素片辞書２０６は全種類のdiphoneを用意しておく必要がある。これに対して、半diphoneを単位とした場合は、足りない半diphoneを別の半diphoneで代替できる。例えば、半diphoneの「/a.b.0/(diphone a.bの左側）」の代わりに「/a.n.0/」を利用しても、音質の劣化を少なくして良好に音声を再生できる。これにより、素片辞書２０６のサイズをより小さくできる。

【００４９】

［実施の形態３］

前述の実施の形態１、２では、音韻の単位としてdiphoneや音素や半diphoneを用いる場合について説明したが、本発明はこれに限定されるものではなく、これらを混合して用いてもよい。例えば、利用頻度が高い音韻については、diphoneを単位とし、利用頻度が低い音韻については、２つの半diphoneを用いて表現するようにしても良い。

【００５０】

図１０は、音声素片単位を混合した場合の一例を示した図で、ここでは音韻「

o.w」がdiphoneで表され、その前後の音韻は半diphoneで表されている。

【００５１】

［実施の形態４］

実施の形態３において、元のデータベース中で連続する場所から取り出されたかどうかの情報を持ち、連続していた場合は、半diphoneの組を仮想的にdiphoneとして扱うようにしてもよい。つまり、元のデータベース中で連続するということは接続歪が“０”であるため、この場合には変形歪だけを考慮すればよいことになり計算量を大幅に軽減できる。

【００５２】

図１１は、この様子を表わした概念図である。図中の線上の数字は接続歪を表している。

【００５３】

図１１において、１１００で示される半diphoneの組は、元のデータベース中で連続する場所から取り出されたものであり、その接続歪みは“０”に一義的に決定されている。また１１０１で示された半diphoneの組は、元のデータベース中で連続する場所から取り出されたものではないため、それぞれに対して接続歪みが計算される。

【００５４】

［実施の形態５］

前述の実施の形態１では，１単位のテキストデータから得られた音韻系列全体を歪計算の対象とする場合について説明したが、本発明はこれに限定されるものでない。例えば、ポーズや無音部分までを一つの区間として音韻系列を分割し、各区間ごとに歪計算を行ってもよい。ここで言う無音部分とは、例えばp,t,kなどの無音部分のことである。ポーズや無音部分では接続歪が“０”であると考えられるため、このような分割が有効となる。これにより、各区間毎に最適な音声素片の選択が可能となる。

【００５５】

［実施の形態６］

前述の実施の形態１では、接続歪の計算にケプストラムを用いる場合について

説明したが、本発明はこれに限定されるものではない。例えば、接続点の前後に亙る波形の差分の和を用いて接続歪を求めても良い。またスペクトル距離などを用いて接続歪を求めてもよい。この場合、接続点はピッチマークに同期させるのが、より好ましい。

【００５６】

［実施の形態７］

前述の実施の形態１では、接続歪の計算において、窓長、シフト長、ケプストラムの次数、フレーム数などを具体的数字を使って説明したが、本発明はこれに限定されるものではない。任意の窓長、シフト長、次数、フレーム数を使って接続歪を算出してもよい。

【００５７】

［実施の形態８］

前述の実施の形態１では、接続歪の計算にケプストラムの次数ごとに差分を取ったものの総和を用いる場合について説明したが、本発明はこれに限定されるものではない。例えば、各次数を統計的性質などを使って正規化（正規化係数ｒｊ）してもよい。この場合の接続歪Ｄｃは、

$$Dc=\Sigma\Sigma\ (rj\times|Cpre\ i,j\ -\ Ccur\ i,j|)$$

となる。ここで、最初のΣはｉ＝－２～２の総和を、次のΣはｊ＝０～１６までの総和を示す。

【００５８】

［実施の形態９］

実施の形態１では、ケプストラムの次数ごとの差分の絶対値をベースに接続歪の算出を行なう場合について説明したが、本発明はこれに限定されるものではない。例えば、差分の絶対値の累乗（累数が偶数の場合は絶対値でなくてもよい）をベースに接続歪の算出を行なってもよい。ここで累数をＮとすると、接続歪Ｄｃは、

$$Dc=\Sigma\Sigma\ |Cpre\ i,j\ -\ Ccur\ i,j|\hat{}N$$

となる。ここで“^Ｎ”はＮの累乗を示す。ここでＮの値を大きくすることは、大きな差分について敏感になることを意味しているので、その結果、接続歪が平均

的に小さくなるように働くことになる。

【００５９】

［実施の形態１０］

前述の実施の形態１では、変形歪としてケプストラムを用いる場合について説明したが、本発明はこれに限定されるものではない。例えば、変形前後の一定区間の波形の差分の和を用いて変形歪を求めてもよい。また、スペクトル距離などを用いて変形歪を求めてもよい。

【００６０】

［実施の形態１１］

前述の実施の形態１では、変形歪を波形から得られる情報を基に算出する場合について説明したが、本発明はこれに限定されるものではない。例えば、PSOLAによるピッチ素片の削除および複製の回数などを変形歪を算出する要素としても良い。

【００６１】

［実施の形態１２］

前述の実施の形態１では、音声素片を読み出すごとに接続歪を計算する場合について説明したが、本発明はこれに限定されるものではない。例えば、接続歪を予め計算しておき、テーブル化して保持してもよいものとする。

【００６２】

図１２は、diphone「/a.r/」とdiphone「/r.i/」との間の接続歪を記憶したテーブルの一例を示す図である。ここでは縦軸に「/a.r/」の音声素片、横軸に「/r.i/」の音声素片をとっている。例えば、「/a.r/」の「id3」の音声素片と「/r.i/」の「id2」の音声素片との接続歪は“３．６”で表されている。このように接続可能なdiphone間の接続歪を全てテーブル化して用意することにより、音声素片同士の合成時の接続歪の算出がテーブルの参照だけで済むため、その計算量を大幅に軽減でき、算出時間を大幅に短縮できる。

【００６３】

［実施の形態１３］

前述の実施の形態１では、音声素片を編集する毎に変形歪を計算する場合につ

いて説明したが、本発明はこれに限定されるものではない。例えば、変形歪を予め計算しておき、テーブルとして保持しておいても良い。

【００６４】

図１３は、あるdiphoneを基本周波数と音韻時間長について変化させた場合の変形歪をテーブルで表した図である。

【００６５】

図中、μは、そのdiphoneの統計的な平均値を示し、σは標準偏差である。具体的な表の作成方法としては、次のような作成方法が考えられる。まず、基本周波数と音韻時間長に関して統計的に平均値と分散を求める。次に、それらを基に（５×５＝）２５通りの基本周波数と音韻時間長をターゲットとしてPSOLA法をそれぞれ適用し、テーブルの変形歪を一つずつ求めていけばよい。合成時は、ターゲットの基本周波数と音韻時間長が決まれば、テーブルの近傍の値で内挿（もしくは外挿）することによって、変形歪を推定することが可能である。

【００６６】

図１４は、合成時に変形歪を推定するための具体例を示した図である。

【００６７】

図中、黒丸がターゲットの基本周波数と音韻時間長であり、このとき、各格子点の変形歪がテーブルからＡ，Ｂ，Ｃ，Ｄと求まっていると仮定すると、変形歪Ｄｔは、以下の式により求めることができる。

$$Dt=\{A\cdot(1-y)+C\cdot y\}\times(1-x)+\{B\cdot(1-y)+D\cdot y\}\times x$$

【００６８】

［実施の形態１４］

前述の実施の形態１３では、変形歪テーブルの格子点として、そのdiphoneの統計的な平均値と標準偏差を基に５×５のテーブルを作成したが、本発明はこれに限定されるものではなく、任意の格子点を持つテーブルとしてもよい。また、格子点を平均値などに依らず決定的に与えてもよいものとする。例えば、韻律推定で推定されうる範囲を等分割するなどもよいものとする。

【００６９】

［実施の形態１５］

前述の実施の形態１では、接続歪と変形歪の重み和で歪を定量化する場合について説明したが本発明はこれに限定されるものではなく、接続歪と変形歪それぞれに閾値を設定しておき、どちらか一方でもその閾値を越えた場合はその音声素片が選択されないようにして、十分大きな歪の値を与えるようにしてもよい。

【００７０】

上記実施の形態においては、各部を同一の計算機上で構成する場合について説明したが本発明はこれに限定されるものではなく、例えばネットワーク上に分散した計算機や処理装置などに分かれて各部を構成してもよい。

【００７１】

上記実施の形態においては、プログラムを制御メモリ（ＲＯＭ）に保持する場合について説明したが本発明はこれに限定されるものではなく、外部記憶など任意の記憶媒体を用いて実現してもよい。また、同様の動作をする回路で実現してもよい。

【００７２】

なお本発明は、複数の機器から構成されるシステムに適用しても、１つの機器からなる装置に適用してもよい。前述した実施の形態の機能を実現するソフトウエアのプログラムコードを記録した記録媒体を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ（またはＣＰＵやＭＰＵ）が記録媒体に格納されたプログラムコードを読み出し実行することによっても達成される。

【００７３】

この場合、記録媒体から読み出されたプログラムコード自体が前述した実施の形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。プログラムコードを供給するための記録媒体としては、例えば、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク，ＣＤ－ＲＯＭ，ＣＤ－Ｒ，磁気テープ、不揮発性のメモリカード、ＲＯＭなどを用いることができる。

【００７４】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施の形態の機能が実現されるだけでなく、そのプログラムコードの指示

に基づき、コンピュータ上で稼働しているＯＳなどが実際の処理の一部または全部を行ない、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

【００７５】

更に、記録媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるＣＰＵなどが実際の処理の一部または全部を行ない、その処理によって前述した実施の形態の機能が実現される場合も含まれるものとする。

【００７６】

以上説明したように本実施の形態によれば、接続歪と変形歪を考慮して素片辞書に登録する音声素片を選択することにより、少数の音声素片を登録した辞書を用いても、音質の劣化が少ない合成音声を生成できるという効果がある。

【００７７】

【発明の効果】

以上説明したように本発明によれば、接続歪や変形歪に基づく歪の影響を考慮して素片辞書に登録する音声素片を選択することによって、そのような素片辞書を用いた合成音声の質を向上できるという効果がある。

【００７８】

また本発明によれば、素片辞書に登録する音声素片の数を少なく抑えて、かつその素片辞書を用いて良好な音声を再生できるという効果がある。

【図面の簡単な説明】

【図１】

本発明の実施の形態に係る音声情報処理装置のハードウェア構成を示すブロック図である。

【図２】

本発明の実施の形態１に係る音声情報処理装置のモジュール構成を示すブロック図である。

【図３】

本実施の形態に係るオンラインモジュールにおける処理の流れを示すフローチャートである。

【図４】

本実施の形態に係るオフラインモジュールの詳細な構成を示すブロック図である。

【図５】

本実施の形態１に係るオフラインモジュールにおける処理の流れを示すフローチャートである。

【図６】

本発明の実施の形態に係る音声素片の変形を説明する図である。

【図７】

本発明の実施の形態に係る音声素片の接続歪を説明する図である。

【図８】

音声素片における歪の決定過程を説明する図である。

【図９】

Ｎbestによる決定過程を説明する図である。

【図１０】

本発明の実施の形態３に係る音声素片の単位をdiphoneと半diphoneとで混合した場合を説明する図である。

【図１１】

本発明の実施の形態４に係る音声素片の単位を取り出した半diphoneによって混合した例を示した図である。

【図１２】

本発明の実施の形態１２に係るdiphoneの /a.r/ と/r.i/間の接続歪を決定するテーブル構成例を示す図である。

【図１３】

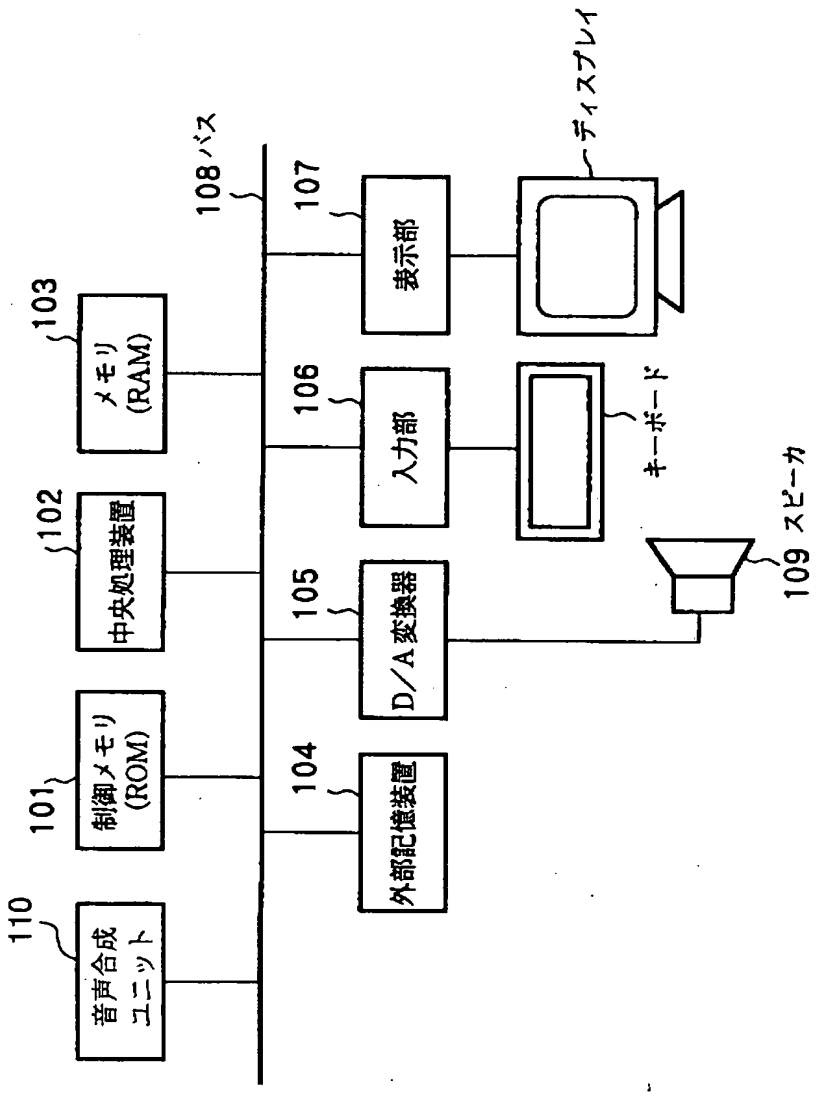
本発明の実施の形態１３に係る変形歪を表わすテーブル例を示す図である。

【図１４】

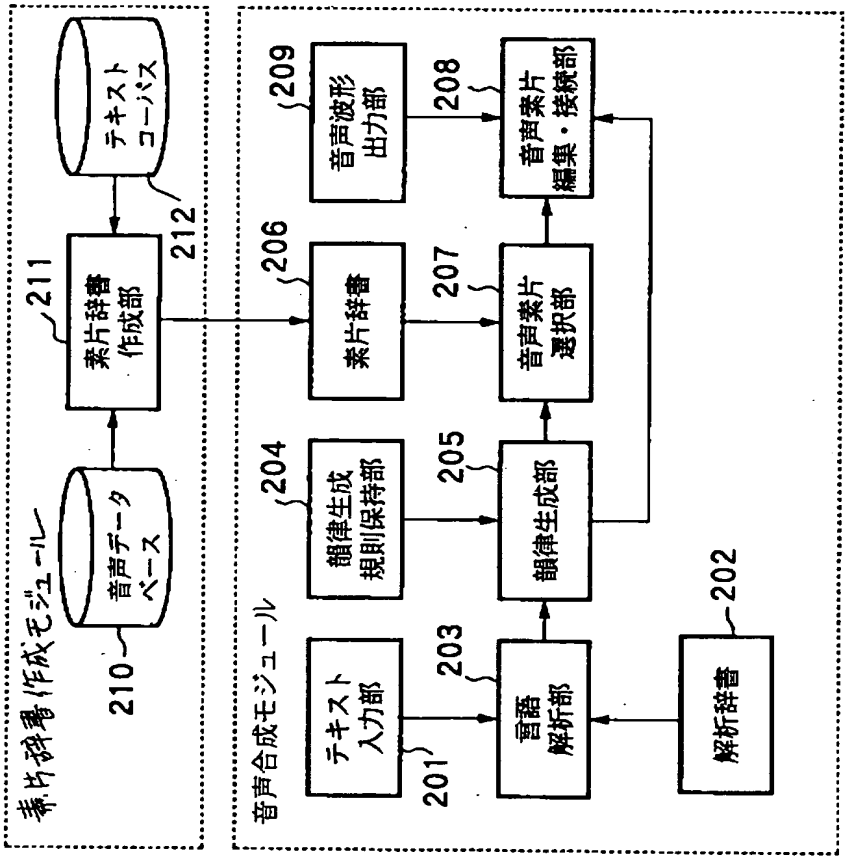
本発明の実施の形態１３に係る変形歪を推定する具体例を示した図である。

【書類名】　図面

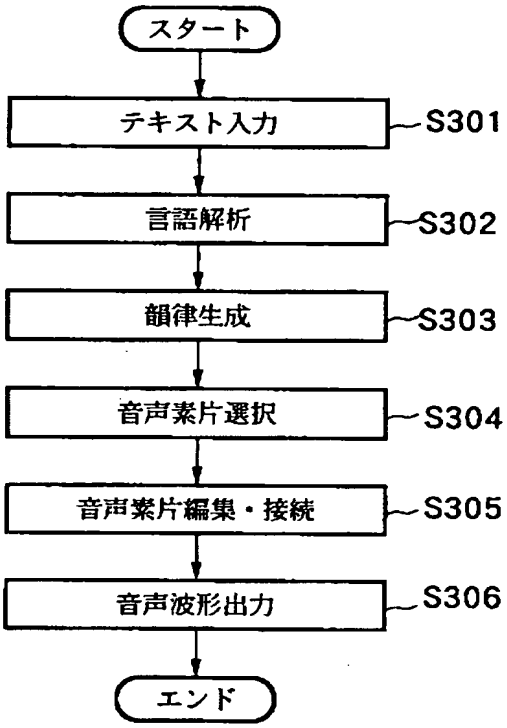
【図１】


ディスプレイ
108 バス
107
表示部
103
メモリ
(RAM)
106
入力部
キーボード
102
中央処理装置
105
D／A変換器
109 スピーカ
101
制御メモリ
(ROM)
104
外部記憶装置
110
音声合成
ユニット

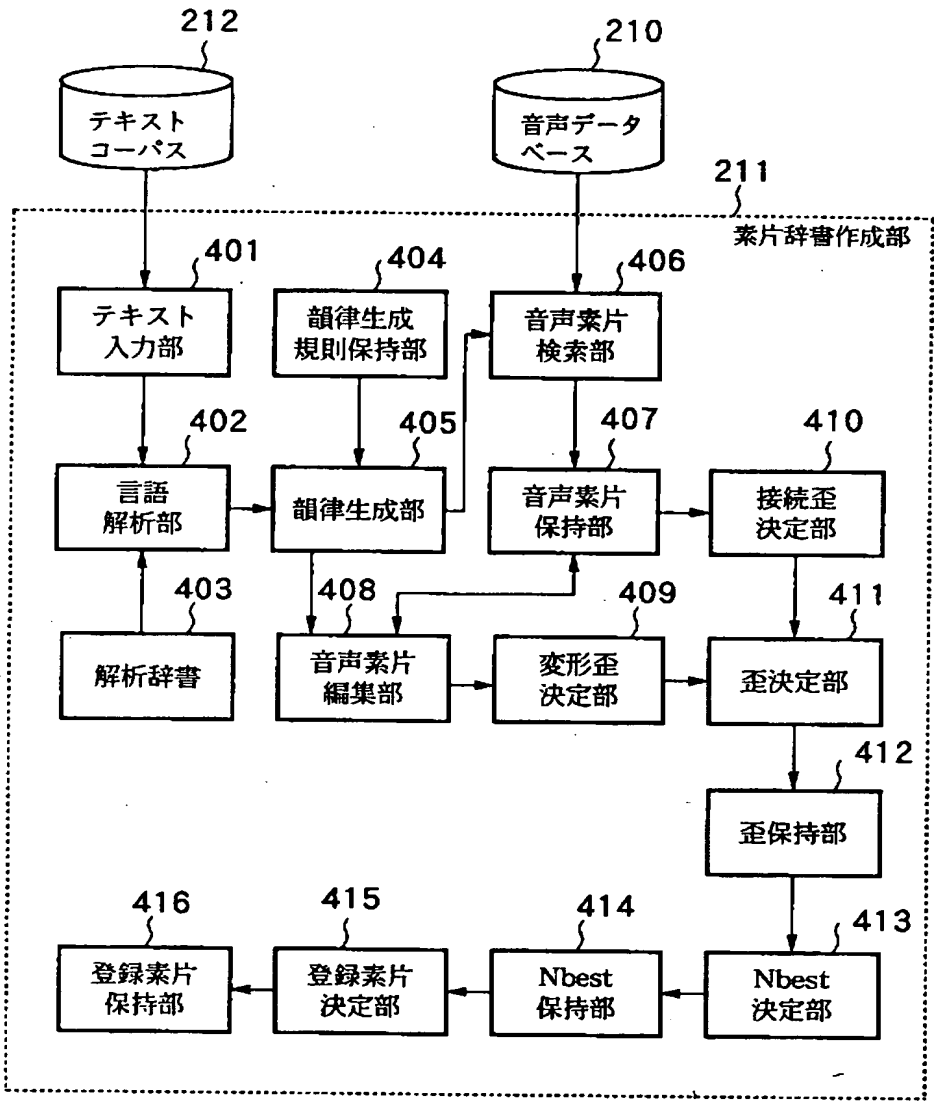
【図２】


素片辞書作成モジュール
音声データベース
210
素片辞書作成部
211
テキストコーパス
212
音声合成モジュール
テキスト入力部
201
言語解析部
203
解析辞書
202
韻律生成規則保持部
204
韻律生成部
205
素片辞書
206
音声素片選択部
207
音声波形出力部
209
音声素片編集・接続部
208

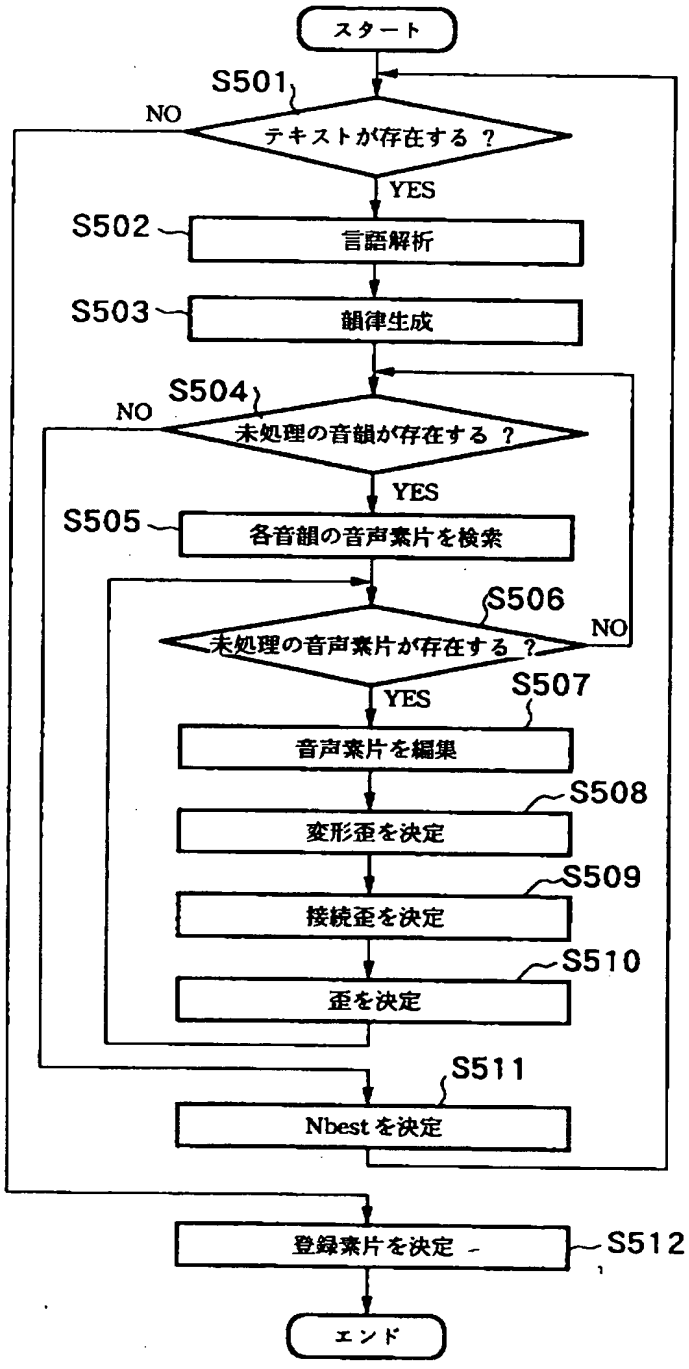
【図３】


スタート
テキスト入力
S301
言語解析
S302
韻律生成
S303
音声素片選択
S304
音声素片編集・接続
S305
音声波形出力
S306
エンド

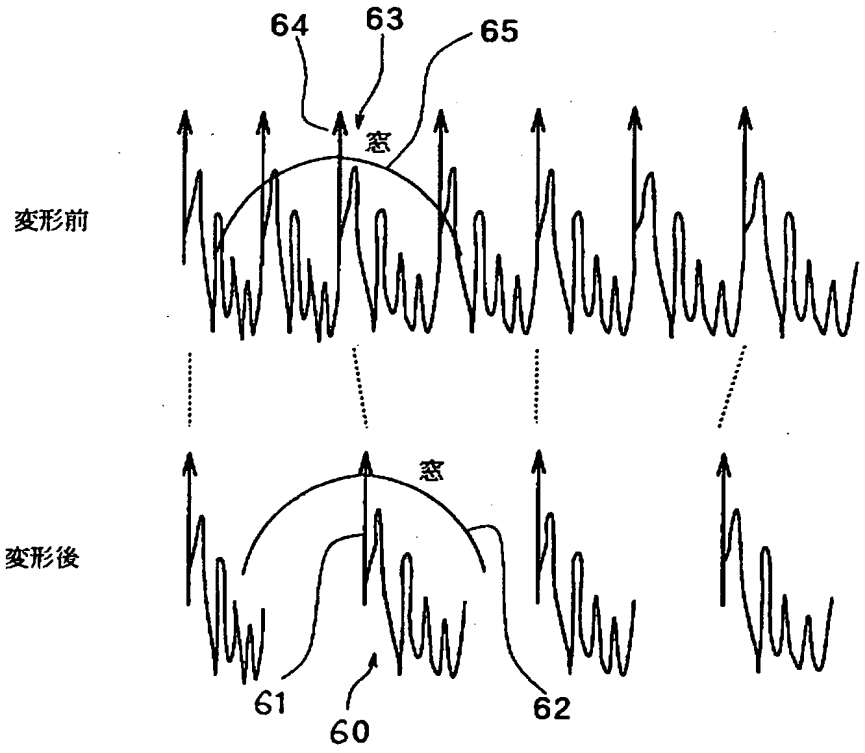
【図４】


212
テキスト
コーパス
210
音声データ
ベース
211
素片辞書作成部
401
テキスト
入力部
404
韻律生成
規則保持部
406
音声素片
検索部
402
言語
解析部
405
韻律生成部
407
音声素片
保持部
410
接続歪
決定部
403
解析辞書
408
音声素片
編集部
409
変形歪
決定部
411
歪決定部
412
歪保持部
416
登録素片
保持部
415
登録素片
決定部
414
Nbest
保持部
413
Nbest
決定部

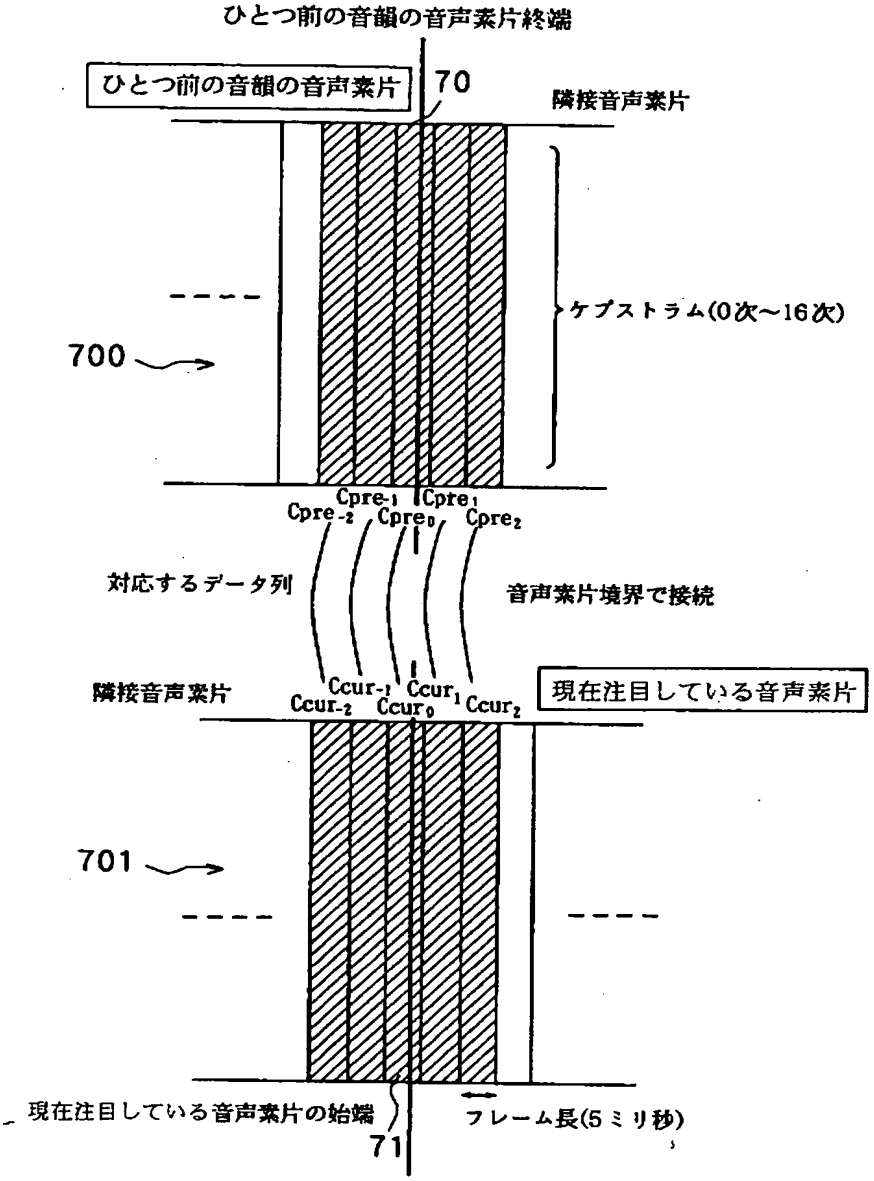
【図５】


スタート
S501
NO
テキストが存在する ？
YES
S502
言語解析
S503
韻律生成
S504
NO
未処理の音韻が存在する ？
YES
S505
各音韻の音声素片を検索
S506
未処理の音声素片が存在する ？
NO
YES
S507
音声素片を編集
S508
変形歪を決定
S509
接続歪を決定
S510
歪を決定
S511
Nbest を決定
登録素片を決定
S512
エンド

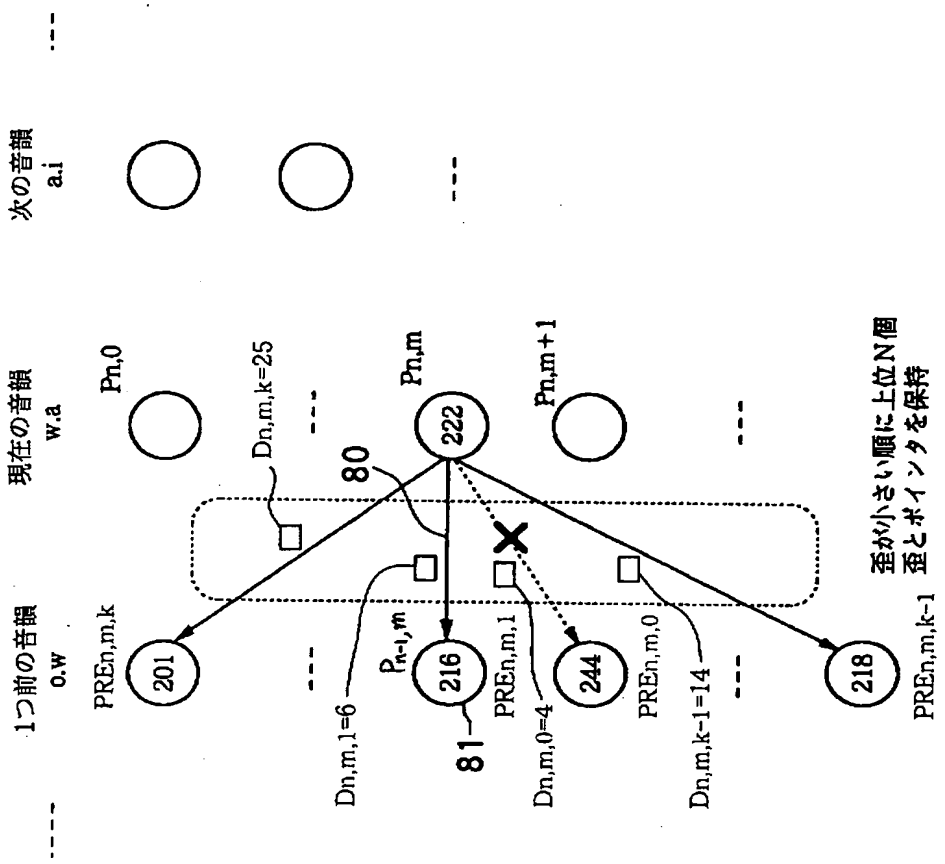
【図６】



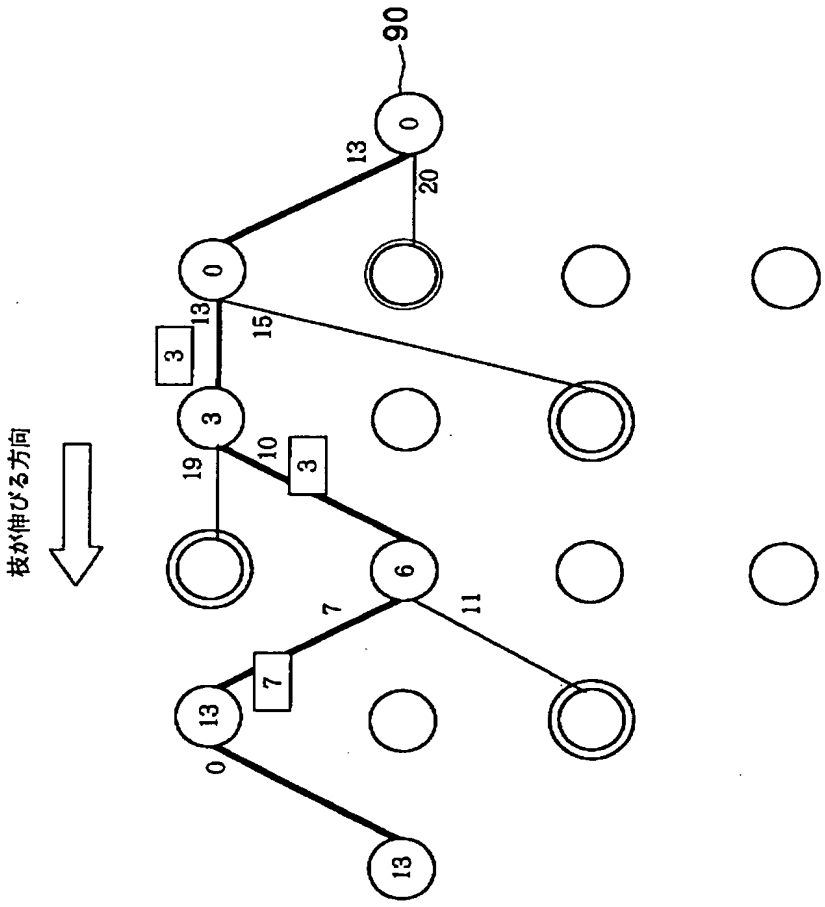
【図７】


ひとつ前の音韻の音声素片終端
ひとつ前の音韻の音声素片
70
隣接音声素片
ケプストラム(0次～16次)
700
Cpre-1
Cpre1
Cpre-2
Cpre0
Cpre2
対応するデータ列
音声素片境界で接続
隣接音声素片
Ccur-1
Ccur1
Ccur-2
Ccur0
Ccur2
現在注目している音声素片
701
現在注目している音声素片の始端
71
フレーム長(5ミリ秒)

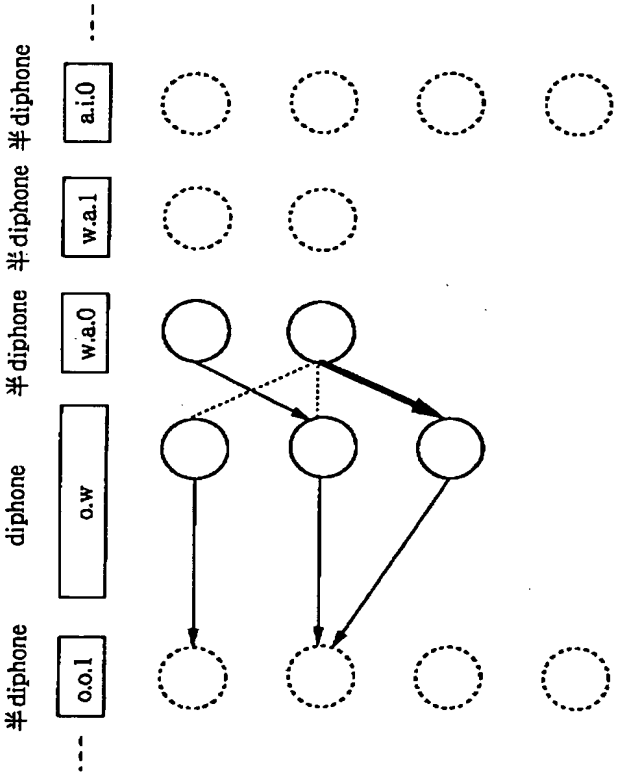
【図８】


次の音韻
a.i
現在の音韻
w.a
1つ前の音韻
o.w
Pn,0
Pn,m
Pn,m+1
Dn,m,k=25
80
81
PREn,m,k
201
Dn,m,1=6
Rn-1,m
216
PREn,m,1
Dn,m,0=4
244
PREn,m,0
Dn,m,k-1=14
218
PREn,m,k-1
歪が小さい順に上位N個
歪とポインタを保持

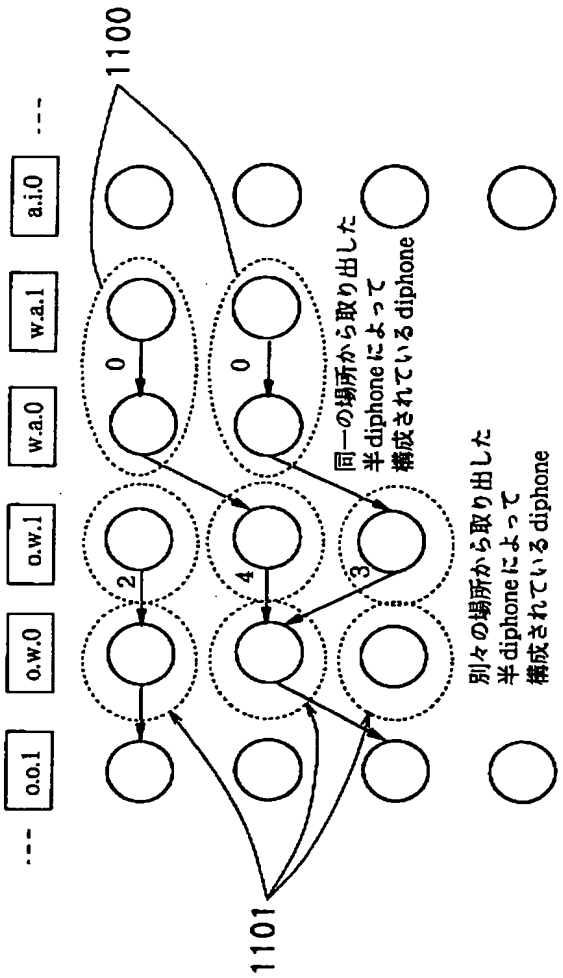
【図９】



【図１０】


半diphone
a.i.0
半diphone
w.a.1
半diphone
w.a.0
diphone
a.w
半diphone
o.o.1

【図１１】


1100
1101
o.o.1
o.w.0
o.w.1
w.a.0
w.a.1
a.i.0
0
0
2
4
3
同一の場所から取り出した
半diphoneによって
構成されているdiphone
別々の場所から取り出した
半diphoneによって
構成されているdiphone

【図１２】

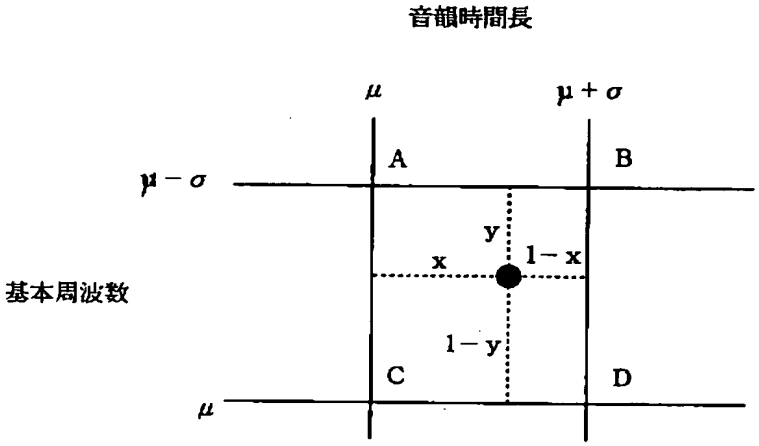
r.i / a.r	id1	id2	id3	id4	id5	...
id1	13.2	2.6	...			
id2	22	11.2				
id3	5	3.6				
id4	0	4				
id5	3.3	...				
...						

【図１３】

音韻時間長

基本周波数		$\mu-2\sigma$	$\mu-\sigma$	μ	$\mu+\sigma$	$\mu+2\sigma$
	$\mu-2\sigma$	26	25	15	13	14
	$\mu-\sigma$	20	24	19	18	17
	μ	15	19	10	13	20
	$\mu+\sigma$	19	22	16	22	26
	$\mu+2\sigma$	25	27	26	30	31

【図１４】



【書類名】　要約書

【要約】

【課題】　素片辞書に登録する音声素片の数を少なく抑えて、かつその素片辞書を用いて良好な音声を再生する。

【解決手段】　入力したテキストデータを言語解析して韻律を生成し、その韻律に基づいて音声データベース２１０から音声素片を検索する。この検索された音声素片の変形歪、及び一つ前の音韻の音声素片との接続による接続歪を求め、歪決定部４１１により、変形歪と接続歪の重み付け等を行なってトータルの歪を決定する。次にNbest決定部４１３により、A*（エースター）探索アルゴリズムを用いて歪が最小となる上位N通りの最適パスを求め、登録素片決定部４１５は上位N通りの最適パスから、その頻度順に素片辞書２０６に登録する登録素片を選び出し、それを素片辞書に登録する。

【選択図】　図４

出　願　人　履　歴　情　報

識別番号　　〔０００００１００７〕

１．変更年月日　　１９９０年　８月３０日
［変更理由］　新規登録
住　所　東京都大田区下丸子３丁目３０番２号
氏　名　キヤノン株式会社